

## IS SOCIAL VALIDITY WHAT WE ARE INTERESTED IN? ARGUMENT FOR A FUNCTIONAL APPROACH

ROBERT P. HAWKINS

WEST VIRGINIA UNIVERSITY

It is argued that neither the term *social* nor the term *validity* is best to identify the processes used or the results obtained in questioning consumers about the goals set, procedures employed, or outcomes achieved in habilitative programming. The term *consumer satisfaction* acknowledges the fact that it is essentially a collection of consumer opinions. The underlying intent of the process might be called *habilitative validation*, a name that seems to better guide our validation efforts. More important, in carefully considering consumer satisfaction assessment, it becomes clear that not only does consumer satisfaction itself need to be validated, but also that more objective methods can be used for assessing habilitative validity. However, legitimate uses still remain for consumer satisfaction measurement, as long as we do not mistake it for strong evidence of the habilitative validity of our goals, procedures, or outcomes.

DESCRIPTORS: consumer satisfaction, habilitative validity, social validity

For over a decade, behavior analysts have been measuring what Wolf (1978) called the "social validity" of the goals they or others set for various learners, the procedures used to reach those goals, or the outcomes actually achieved. This measurement is typically done by questioning either the direct consumer (the learner) or some indirect consumer—family members, teachers, community members, or referring agents—by asking questions such as how satisfied they are with the goals, procedures, or outcomes. I will attempt to clarify what is being measured by such questions, what this implies regarding our name for the process, the relationship between such measurement and behavioral programming, and implications for our research.

### *What is Being Measured in Social Validation, and What Should We Call the Process?*

Asking consumers how satisfied they are with a goal, procedure, or outcome is, of course, measuring verbal behavior. It is essentially the same as asking for an opinion or judgment. Asking a consumer

"Is this an important goal to achieve?" or "Did our services help your son?" is asking for an opinion. Even asking "How acceptable is this procedure?" seems essentially the same as asking for an opinion in that it seems functionally equivalent to asking "Will this procedure probably work?" or "Do you think I know what I am doing?"

Calling this a measurement of social validity seems misleading, for two reasons. The first is perhaps of minor importance but is worth mentioning: The word "social" is likely to imply that the goals, procedures, or outcomes are social ones. Thus "social validity" sounds like "validity for society" or "validity for interpersonal interactions." Yet many goals and outcomes are personal, not social or interpersonal. A child's learning to put on his or her shoes, a physically handicapped person's learning to prepare meals, a depressed person's learning to phone acquaintances and invite them over, or a poor person's learning to shop wisely are primarily for their own personal benefit, not society's. Nor do they have much to do with interpersonal interactions. Although applied behavior analysis has taken an unusual and healthy interest in societal problems (Krasner, 1988), using the term *social validation* implies that those are our only interests.

One might argue that the term *social* does not refer to the type of validity being assessed but rather to the process used: asking various persons (society?). Thus, the validation process is a social one;

This paper is based partly on a paper by the same title presented at the Association for Behavior Analysis convention, Nashville, Tennessee, 1990.

Thanks to Wayne Fuqua for comments on an earlier version of this manuscript. Direct reprint requests to Robert P. Hawkins, Department of Psychology, West Virginia University, Morgantown, West Virginia 26506-6040.

but we should recognize that the validity being assessed is not necessarily social, so the term *social validity* is misleading. A more accurate term for the process of gathering consumer opinions seems to be one used in a special 1983 issue of *Behavior Therapy*: consumer satisfaction.

### *Relation Between Consumer Satisfaction and Habilitation*

A more serious problem is presented by the second word in "social validation." The validity of any test, observation, or other assessment device might be defined as how well the device performs the function it is intended or assumed to perform. The best evaluation of the validity of an assessment device is to get data showing that it does function in the manner intended, as demonstrated by criterion measures more credible than the assessment device itself.

In applied behavior analysis and behavior therapy, researchers often seem to assume that a high satisfaction rating from consumers demonstrates that the goal, procedure, or outcome involved is well chosen. But assessing consumer satisfaction is merely obtaining a second opinion, one to add to the behavioral programmer's own opinion or choice of the goal, outcome, or procedure. A second and concurring opinion is somewhat reassuring, but it does not validate the programmer's choices in the sense of comparing those choices with some more credible criterion. In fact, a consumer's verbal behavior about a goal, procedure, or outcome is often less credible than the professional's own judgment. As Parloff (1983) said, "the patient's expertise in judging personal 'acceptability' of practice and procedures is not an adequate substitute for authority to judge need for services or their appropriateness" (p. 243).

Parloff may underestimate the potential value of measuring consumer satisfaction, but the term *social validity* and some current uses of social validation procedures suggest an overestimation of the value. In fact, rather than using consumer satisfaction measures to validate the programmer's choices of goal, outcome, or procedure, we need to validate the consumer satisfaction ratings them-

selves. I will discuss this further after defining the kind of validity we seem to really be interested in.

*What would validate a goal, outcome, or procedure?* If consumer opinion is not the best criterion for evaluating a goal, procedure, or outcome, what is? I suggest that what we are interested in is the habilitative value or habilitative validity of that goal, procedure, or outcome. Habilitative validity is consistent with a behavior-analytic perspective and should be the criterion in evaluating the validity of any assessment device, including tests. Defining a test's validity as "the degree to which it measures what it purports to measure" is sufficient only if you take a mentalistic viewpoint.

To understand the concept of habilitative validity, it helps to first define habilitation or adjustment. I define habilitation as the degree to which the individual behavioral repertoire maximizes the overall benefits and minimizes the overall costs to that individual and to others, including family, peers, and society (Hawkins, 1986). The benefits and costs to be considered include both short-term and long-term ones. The short-term benefits and costs in the person's natural environment tend to reinforce or punish behavior effectively. Thus, the repertoire already reflects these through the processes Skinner (1974) called contingency shaping. However, long-term benefits and costs either fail to exert control over the person's behavior or do so only indirectly and crudely through the "rules" (Skinner, 1974) and practices of the culture or individual. Therefore, it behooves the programmer who is attempting to assess habilitative validity to emphasize consideration of the long-term natural consequences of various behavior patterns. This consideration includes benefits and costs with low probability but large magnitude, such as the risk of contracting AIDS or the chance of getting a job for which there are dozens of other applicants.

To the extent that a consumer's reported satisfaction predicts the benefits and costs delivered by the natural environment as a result of the particular goals set, outcomes achieved, or procedures used—especially the longer term benefits and costs—that report has habilitative validity. It predicts the habilitative value of the goal, outcome, or procedure.

Unfortunately, many researchers seem to conceive of consumer satisfaction as the ultimate criterion rather than an opinion that may or may not predict the real criteria; studies that actually test the habilitative validity of consumer satisfaction reports seem rare.

*Testing the habilitative validity of consumer satisfaction: predicting the consumer's own behavior.* Studying the habilitative validity of consumer satisfaction reports could take many forms. One type of study would test the ability of such reports to predict the behavior of the very consumers providing them (Fuqua & Schwade, 1986; Hawkins, Conaway, & Conaway, 1986). When Kazdin (1980) wrote "Treatments viewed by the public as more acceptable than others are more likely to be sought by potential consumers, initiated, and adhered to once they are initiated . . ." (p. 260), he made it clear that one reason for obtaining consumer judgments is to predict that same consumer's behavior: his or her seeking, cooperating with, and persistently implementing behavioral programming. Kazdin (1980) used the term *treatment acceptability* for a consumer's judgment about a procedure, thus emphasizing such predictive purpose. This predictive value seems to have been a major interest of Wolf (1978) as well. Unfortunately, from Calvert and Johnston's (1990) recent review of the research on treatment acceptability, it appears that none of that research addressed the validity of consumers' acceptability ratings in predicting either the respondents' behavior or any other events. Instead, the studies investigated how certain variables affected treatment acceptability, as though the validity of treatment acceptability were already established. Other limitations of that research have been discussed by Fuqua and Schwade (1986).

A study by Barlow, Reynolds, and Agras (1973) illustrates the limited ability of verbal reports of treatment acceptability to predict the consumers' own behaviors, even their subsequent verbal reports. Although Barlow et al. did not measure treatment acceptability formally, the study is still relevant. A young man came to them who exhibited extensive transsexual behavior. He requested a sex-

change operation. Probably his long history of very feminine behavior made it seem unlikely to him that he could be taught to behave in a masculine manner. Only reluctantly and tentatively did he accept the researchers' goal of augmenting the masculinity of his behavior and accept their intended procedures. However, they pursued their goal and used their procedures, with the result of greatly increasing the masculinity of his behavior. They reported he was very satisfied with the outcome and was no longer interested in an operation. Thus, the young man's initial acceptance of the treatment was a poor predictor of his own later acceptance (and of the treatment's effectiveness), whereas the professionals' opinion of the treatment initially was an excellent predictor of their later acceptance and even the client's acceptance (and the treatment's effectiveness). The consumer's opinion, in this case, would have been a poor criterion for validating the professionals' decisions or for selecting either goals or procedures.

Most clinical behavior analysts have experienced many cases in which the consumer did not accept a procedure, even though the clinician was certain of its habilitative value. Further, most clinicians could describe cases in which they engaged in extensive persuasion before achieving even moderate client acceptance of a procedure, or cases in which their persuasion failed. It seems clear that the habilitative validity of consumer acceptance cannot be assumed and needs to be evaluated. We need to know, for example, how well a parent's reported satisfaction with the goals set for their child (or for the parent) predicts their return for treatment, their conscientious implementation of data gathering and intervention, their paying for the service promptly, their recommending the professional to others, and so on. Similarly, we need to know how well a psychotic adult's acceptance of a proposed procedure predicts his or her cooperation with that procedure, recording of data, complaints to others, and so forth.

*Testing the habilitative validity of consumers' satisfaction: predicting events other than their own behavior.* Besides predicting their own behavior, we expect consumer satisfaction reports to

predict such things as the habilitative value of goals, the adequacy of outcomes, or the effectiveness of procedures to be used. Kirigin, Braukmann, Atwater, and Wolf (1982) conducted one such predictive validity study when they correlated youth satisfaction with records reflecting treatment effectiveness. They found high predictive validity, but it is easy to find cases in which consumer judgments were poor predictors of the actual benefits or costs of achieving a certain goal through a certain procedure.

I once knew a young retarded man whose teacher noticed that his teeth were not being well cared for. The teacher suggested to his parents that one appropriate goal for his individual education plan would be for him to learn to brush his teeth, a skill he could have acquired within a few weeks. The parents said it was not worth the effort and instead had all his teeth pulled out. By such a decision, I believe his parents were risking creation of a more difficult teaching problem (wearing his dentures) and probably diminishing his acceptability in the community. Thus they were risking the cost of more difficult teaching and jeopardizing many future benefits arising from routine involvement in social relations. Again, the consumer's opinion is not a sufficient criterion to be called a validation; it is merely a second opinion from a different viewpoint.

Many other examples could be given to illustrate low validity of consumer judgments; but it would be a mistake to imply that all behavioral programmers' judgments are necessarily valid, as demonstrated in Winett and Winkler's (1972) criticism of educators and behavior analysts for emphasizing quiet, docile behavior among school children (see also Emery & Marholin, 1977). It would also be a mistake to suggest that consumer satisfaction ratings have no validity. Instead we should recognize that both the professional's and the consumer's judgments are likely to be wrong part of the time and right part of the time.

My point here is that until studies are conducted to discover what sorts of habilitative validity we can expect of consumer ratings and under what circumstances, it seems misguided to act as though consumer satisfaction reports have such validity.

Their value thus far has been one of providing a comforting confirmation of behavioral programmers' opinion, not necessarily an accurate prediction of benefits or costs. Further, our acting as though consumer satisfaction reports validate our goals, outcomes, or procedures may impede our use of more credible and objective methods of validating the habilitative validity of our goals, outcomes, and procedures.

Before discussing objective methods of habilitative validation, I would like to identify a potential problem in our assessment of consumer satisfaction as a prediction of either the consumer's own behavior or any other events: the use of global and ambiguous questions. When we ask a consumer "How satisfied are you with the appropriateness of this goal?," the consumer is likely to respond as though the question were "Have I given you enough chance to talk about the problem?," "How difficult do you think this goal will be to achieve?," or "How well do you like me?" Answers to such questions are unlikely to be good predictors of any specific class of events and thus will be difficult to test for validity in any credible way. Thus, one improvement we might make in assessing consumer satisfaction is to think of what we want a satisfaction rating to predict, then direct our question at that prediction, like this: "Imagine it is a year from now and this goal has already been achieved. How much better is \_\_\_\_\_'s life, in terms of what she can do, how easy life is for her, how many pleasant things happen to her, and so on?" (Hawkins et al., 1986).

### *Objective Methods for Validating Our Goals, Outcomes, or Procedures*

Any educational, therapeutic, preventive, performance-enhancing, or performance-maintaining efforts that affect the behavior of one or more persons in ways that increase the probability or magnitude of benefits (in a sense, reinforcers) for them or others, or that decrease the probability or magnitude of various costs (in a sense, punishers) are habilitative (Hawkins, 1986). Thus, teaching children to read, solve mathematical problems, or make friends with peers is likely to be habilitative, because

these efforts generally increase benefits and decrease costs for both the individual learner and the rest of society. Similarly, teaching adults to use safety belts, to refuse harmful drugs, or to interact positively with their spouse or offspring is probably habilitative, again because it improves the probable benefit-to-cost ratio.

Using this concept, we have some notion of what variables constitute relevant habilitative validity criteria. A goal, outcome, or procedure is valid to the extent that choosing it actually improves the benefit-to-cost ratio for the individual or for others, or both. A consumer's or professional's opinion about a goal, outcome, or procedure is only valid to the extent that it is consistent with such improved benefit-to-cost ratio.

The following discussion of objective methods is not meant to imply that benefits and costs can, themselves, always be objectively identified and quantified; at some point in a comprehensive validation, subjective judgment will likely be needed. But using subjective judgment as the first and only criterion ignores the objective possibilities.

*Studying average or exemplary performance.* One objective validation strategy is the use of norms or epidemiological data to estimate what behaviors are likely to be habilitative (see Winett, Moore, & Anderson, 1991). Clinicians, educators, developmentalists, and others have used such data for decades in estimating what behavior might be adaptive (see Hawkins, 1975; Kazdin, 1977; Lent, 1968). Performance that is statistically normative can often be taken as adequately adaptive (effective), even though it may be far from ideal. For example, Johnson, Wahl, Martin, and Johansson's (1973) data on the frequency with which youngsters of certain ages wet their beds can be used by a behavior analyst to decide whether to select, as a target behavior, occasional bedwetting in, say, a 6-year-old boy.

One can also measure the behavior of an exemplary group to determine appropriate behaviors to target, appropriate stimulus conditions, or appropriate performance criteria, as suggested by Gilbert (1978) and Van Houten (1979). Van Houten gave the example of a typist's performance and

pointed out cogently that it might be better to select exemplary typists in setting performance criteria rather than average, typical typists. Similarly, one might observe the behavior of highly successful supervisors to see what they do that may account for the cooperation from their employees.

*Studying the association between performance and apparent effectiveness.* A very similar form of research is to correlate specific behaviors with objective indices of adjustment or effectiveness. For example, Holmes, Hansen, and St. Lawrence (1984) used selected comparison groups—comparing the conversational skills of former psychiatric hospital patients with those of normal community persons—and from these data, they selected both the target behaviors and the criteria for adequate performance. A somewhat different correlational strategy is to study a wide range of persons, correlating data on effectiveness with data on specific behaviors under specific stimulus conditions. For example, college men who vary in their success at getting dates could be observed interacting with women to see whether behaviors can be determined to account for their effectiveness at getting dates. Similarly, the correlation between smoking and various diseases suggests the importance of this behavior. Obviously, these methods do not prove that the measured behaviors are functional, but they certainly suggest possibilities.

*Experimental analysis of alternative performances.* The best validation of what behavior is most adaptive is to test experimentally the outcomes produced by different behaviors and different levels of their performance (Fuqua & Schwade, 1986; Hawkins et al., 1986). The performance yielding the greatest benefit at the least cost is the most adaptive, by definition (Hawkins, 1986). This approach was used by Warren, Rogers-Warren, and Baer (1976) when they assessed the effect of different frequencies of children's offers to share materials. They measured peers' reactions of accepting the offer and found that such acceptance was maximal when the target child made offers at a middle frequency, neither very often nor very seldom. This peer reaction was a very relevant measure of the target behavior's adaptiveness, because

it is the immediate reinforcer that should eventually maintain the behavior. Of course other benefits could also have been studied, perhaps longer term ones, such as peers' approaching the child or peers' offering to share. It should be noted that consumer satisfaction ratings by anyone—children, teachers, parents, or child development experts—probably would not have led to so accurate a conclusion.

Jones and Azrin (1969) also did an objective habilitative validation when they varied the speed of a metronome to determine what speed produced the most "natural" speech from stutterers who were pacing their speech with the metronome. Similarly, one could validate the targeting of various preacademic and academic skills by testing the degree to which they facilitated acquisition of subsequent academic skills.

*Experimental analysis of alternative interventions.* Thus far, my analysis of objective strategies of validation has emphasized methods for validating a goal or an outcome. The methods described are common forms of research, but not nearly as common among behavior analysts as is our validating an intervention objectively. The *Journal of Applied Behavior Analysis* is filled with demonstrations that an intervention was effective, that it had more habilitative validity than certain other procedures. Further, when we compare one sophisticated treatment with another—as opposed to comparing a sophisticated treatment with unsophisticated procedures already in use ("no treatment")—we add a very refined form of habilitative validation.

#### *Promising Uses for Consumer Satisfaction Measurement*

Of what value, then, are consumer satisfaction measures? I see at least seven reasons to measure consumer satisfaction, provided we do not delude ourselves about what the resulting data mean.

*To suggest the need for education.* Suppose a behavior analyst suggests that parents apply a home token economy to their child's school performance and ask the parents what they think of that idea. If the parents give a low opinion of the procedure, perhaps saying they object to "bribing" their child,

we would be wise to assume that this verbal behavior is predictive of implementation problems, whether we have tested the validity of that prediction or not. Of course it would not be wise to discard the plan of using a token economy if our own history suggests it will be effective; instead, we would likely engage in educating the parent about the definition of bribery and the universality of reinforcement. Or suppose we ask a hypochondriacal woman if it will be acceptable for her husband and children to walk away whenever she complains of physical problems. If she says this is not acceptable, we may then spend considerable time explaining what we know about operant behavior, such as pointing out that reinforcement is automatic and that, although the complaints are genuine, they cost her and others in the long run. The same analysis applies to consumer acceptance of treatment goals and effects.

Of course consumer education will have its limits. For example, I know a mildly retarded client who is his own guardian and who has repeatedly rejected certain goals and procedures that professionals have tried to convince him would be truly habilitative. Even though the goals and procedures would be both habilitative and humane, the professionals are left with the ethical paradox that the consumer will not accept the treatment, yet denying effective treatment is also problematic. Criminals, alcoholics, and many others will often reject our educational efforts as well, even in cases in which the long-term prognosis is life threatening. In a humane, democratic society such ethical dilemmas are inevitable.

*To discover further behavioral and environmental resources.* The learner and those who are directly affected by his or her behavior have vastly more experience, compared to the professional, at observing the learner's behavior—motor, verbal, and affective (e.g., "feelings")—and the behavior's antecedents and consequences. These consumers also know of behavioral and material resources that could be used in habilitative programming. A behavior analyst's effectiveness would be quite limited if he or she failed to assess such factors, and formal consumer satisfaction measurement can be part of

such assessment. Of course, informal assessment of consumer satisfaction is routine and continuous in human services. The professional frequently asks questions and gives other opportunities for the consumers with whom he or she meets to give opinions about goals, interventions, and outcomes; the sensitive professional notes both *what* the consumer says and the paraverbal behaviors that constitute *how* it is said: intonation, timing, facial expression, body posture, and so forth. This is one reason why interviewing is so universal in all applied work (see Linehan, 1977). Although neither formally nor informally assessed satisfaction should be assumed to have high habilitative validity, it can provide suggestive evidence of untapped resources and should lead to additional, more valid assessment to determine whether those resources are real and valuable.

*To suggest promising adjustments in procedures.* In a foster-family-based treatment program I have been involved with for several years, youngsters often object to having a point system applied to their behavior, even after experiencing it for many months. This constitutes low treatment acceptability. The staff, wisely, would not forego the many advantages of token economies on the basis of such an evaluation by the youngsters, but the evaluations did lead the staff to use the point systems of some youngsters as consequences themselves. Youngsters who were performing very well under a point system were allowed to go without it, as a reinforcer for continuous good performance, with the agreement that it would be reinstated immediately if their behavior deteriorated in certain ways. I do not really know whether this set of contingencies actually produced better performance, but it seems quite possible that it did. If so, it would show that although a consumer's satisfaction may not be taken at face value, it may lead a responsive behavioral programmer to adjustment procedures in ways that have favorable effects.

Sometimes a report of low satisfaction can suggest ways to make a program more humane or palatable to the direct consumer. If a client complains about a procedure, for example, only an insensitive programmer will neglect to reconsider exactly how that procedure is being applied. Often

small adjustments that seem unrelated to achieving the targeted behavior changes can make the procedure much more acceptable to the consumer and occasionally more effective in achieving the targeted changes as well.

*To predict or detect undesired effects.* Many things can go awry in any habilitative program. Predicting such effects often permits one to minimize their probability and severity. Consumers often have information relevant to such predictions, so their report—including their reported satisfaction—can be a valuable predictor. Even when low satisfaction accurately indicates undesired effects only occasionally, it can be useful, because it can lead to other assessments of greater validity and thus facilitate corrective action, if needed.

*To assess the comprehensiveness of the effects.* It is common, at least in clinical services, for a client to begin with a particular set of complaints and yet not be satisfied when those complaints have been addressed successfully. If such dissatisfaction is followed by further assessment that shows there are indeed further problems, assessing the dissatisfaction did have some habilitative validity.

*To document a program's effects on numerous clients.* It is rarely feasible for an ongoing service to measure all of its effects on all of its clients in a highly valid, objective manner. That would be equivalent to making the entire service a form of elaborate research. Yet accountability is very important, especially to a behavior analyst. A compromise is to obtain objective data on at least some effects with at least some clients, while assessing the impact on remaining clients indirectly and economically through consumer satisfaction and similar questionnaires and checklists. The assumption, of course, is that the consumer satisfaction assessment has sufficient validity to be worthy as one criterion of effectiveness, although it is not a sufficient criterion by itself. Because consumer satisfaction can be extremely economical to assess, it can be assessed frequently—quarterly, monthly, or even weekly—providing the behavioral programmer with fairly frequent opportunity to obtain the potential benefits already mentioned.

*For public relations.* Most of us appreciate others' asking our opinions. Thus, in the process of

asking a consumer for opinions, the behavior analyst probably often favorably affects the very opinions being requested. This may increase the probability of other actions that will benefit the behavior analyst, such as the clients' paying their bills or recommending the service.

Further, consumer satisfaction data can be combined with objective data in an overall program evaluation and used to make favorable impressions on boards of directors, funding sources, referral sources, and other indirect consumers. Most ongoing human services collect and use no evaluative data, at least on outcomes (Carter, 1983); therefore, a service that does collect such data stands out. The fact that some of the data are consumer satisfaction measures does not preclude a favorable impression, especially if these subjective measurements are combined with objective data.

### Conclusion

What we are directly assessing when we ask a person how satisfied he or she is with a goal, a treatment, or an outcome is the person's verbal repertoire. Baer, Wolf, and Risley (1968) said, in defining applied behavior analysis, "a subject's verbal description of his own non-verbal behavior usually would not be accepted as a measure of his actual behavior" (p. 93). This value continues to be appropriate in applied behavior analysis and is relevant in evaluating the validity of a goal, procedure, or outcome. In measuring consumers' verbal judgments, we are only hoping that these verbal behaviors are substantially controlled by variables directly relevant to the habilitation task at hand, and thus that they predict habilitative outcomes to some degree. The validity of such consumer judgments has yet to be established; they should not be viewed as a validity criterion but rather as a second opinion from a lay person who may or may not be better informed and less biased than the professional is. If we view consumer satisfaction from this more realistic perspective we can still find many uses for its measurement, using questions designed to facilitate the habilitative validity of the consumers' answers. However, we should make greater use of the many methods for objectively

evaluating the habilitative value of the goals we select, the procedures we devise, and the outcomes we obtain.

### REFERENCES

- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91-97.
- Barlow, D. H., Reynolds, E. G., & Agras, W. S. (1973). Gender identity change in a transsexual. *Archives of General Psychiatry*, 28, 569-579.
- Calvert, S. C., & Johnston, C. (1990). Acceptability of treatments for child behavior problems: Issues and implications for future research. *Journal of Clinical Child Psychology*, 19, 61-74.
- Carter, R. K. (1983). *The accountable agency*. Beverly Hills, CA: Sage.
- Emery, R. E., & Marholin, D., II. (1977). An applied behavior analysis of delinquency: The irrelevancy of relevant behavior. *American Psychologist*, 32, 860-873.
- Fuqua, R. W., & Schwade, J. (1986). Social validation of applied behavioral research: A selective review and critique. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 265-292). New York: Plenum.
- Gilbert, T. F. (1978). *Human competence: Engineering worthy performance*. New York: McGraw-Hill.
- Hawkins, R. P. (1975). Who decided *that* was the problem?: Two stages of responsibility for applied behavior analysts. In W. S. Wood (Ed.), *Issues in evaluating behavior modification* (pp. 195-214).
- Hawkins, R. P. (1986). Selection of target behaviors. In R. O. Nelson & S. C. Hayes (Eds.), *Conceptual foundations of behavioral assessment* (pp. 331-385). New York: Guilford.
- Hawkins, R. P., Conaway, R. L., & Conaway, L. P. (1986, May). What is "social" about social validity? Invited address presented at the Association for Behavior Analysis convention, Milwaukee, WI.
- Holmes, M. R., Hansen, D. J., & St. Lawrence, J. S. (1984). Conversational skill training with aftercare patients in the community: Social validation and generalization. *Behavior Therapy*, 15, 84-100.
- Johnson, S. M., Wahl, G., Martin, S., & Johansson, S. (1973). How deviant is the normal child?: A behavioral analysis of the preschool child and his family. In R. D. Rubin, J. P. Brady, & J. D. Henderson (Eds.), *Advances in behavior therapy* (Vol. 4). New York: Academic.
- Jones, R. J., & Azrin, N. H. (1969). Behavioral engineering: Stuttering as a function of stimulus duration during speech synchronization. *Journal of Applied Behavior Analysis*, 2, 223-229.
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1, 427-452.
- Kazdin, A. E. (1980). Acceptability of alternative treatments for deviant child behavior. *Journal of Applied Behavior Analysis*, 13, 259-273.



- Kirigin, K. A., Braukmann, C. J., Atwater, J. D., & Wolf, M. M. (1982). An evaluation of teaching-family (Achievement Place) group homes for juvenile offenders. *Journal of Applied Behavior Analysis*, *15*, 1-16.
- Krasner, L. (1988). Paradigm lost: On a historical/sociological/economic perspective. In D. B. Fishman, F. Rogers, & C. M. Franks (Eds.), *Paradigms in behavior therapy. Present and promise* (pp. 23-44). New York: Springer.
- Lent, J. R. (1968). Mimosa Cottage: Experiment in hope. *Psychology Today*, *2*, 157ff.
- Linehan, M. M. (1977). Issues in behavioral interviewing. In J. D. Cone & R. P. Hawkins (Eds.), *Behavioral assessment: New directions in clinical psychology* (pp. 30-51). New York: Brunner/Mazel.
- Parloff, M. B. (1983). Who will be satisfied by "consumer satisfaction" evidence? *Behavior Therapy*, *14*, 242-246.
- Skinner, B. F. (1974). *About behaviorism*. New York: Knopf.
- Van Houten, R. (1979). Social validation: The evolution of standards of competency for target behaviors. *Journal of Applied Behavior Analysis*, *12*, 581-591.
- Warren, S. F., Rogers-Warren, A., & Baer, D. M. (1976). The role of offer rates in controlling sharing by young children. *Journal of Applied Behavior Analysis*, *9*, 491-497.
- Winett, R. A., Moore, J. F., & Anderson, E. S. (1991). Extending the concept of social validity: Behavior analysis for disease prevention and health promotion. *Journal of Applied Behavior Analysis*, *24*, 215-230.
- Winett, R. A., & Winkler, R. C. (1972). Current behavior modification in the classroom: Be still, be quiet, be docile. *Journal of Applied Behavior Analysis*, *5*, 499-504.
- Wolf, M. M. (1978). Social validity: The case for subjective measurement, or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*, 315-329.

Received February 3, 1991

Initial editorial decision February 21, 1991

Revision received March 5, 1991

Final acceptance March 8, 1991

Action Editor, E. Scott Geller